

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
КАФЕДРА ТЕОРИИ СИСТЕМ УПРАВЛЕНИЯ ЭЛЕКТРОФИЗИЧЕСКОЙ АППАРАТУРОЙ

Килин Павел Дмитриевич

Выпускная квалификационная работа бакалавра

**Математическая обработка данных ПЭТ
исследований головного мозга**

Направление 010900

Прикладные математика и физика

Научный руководитель,
доктор физ.-мат. наук,
профессор кафедры Теории
управления
Котина Е. Д.

Санкт-Петербург

2017

Содержание

Введение	3
Постановка задачи	5
Глава 1. Модель исследования	6
1.1. Камерные модели в исследованиях головного мозга	6
1.2. Графическая модель Logan	10
1.3. Неинвазивная модель Ichise	13
Глава 2. Математическая обработка данных	14
2.1. Модель наблюдения.....	14
2.2. Регрессия на главные компоненты	16
2.3. Гребневая регрессия и кросс-валидация	18
2.4. Применение методов и анализ результатов	20
Заключение	23
Список литературы	24
Приложение	26

Введение

Позитронная эмиссионная томография (ПЭТ) – это диагностическая процедура исследования внутренних органов с использованием метода радионуклидной диагностики. Информация, полученная с помощью ПЭТ, позволяет выявить патологические процессы на самых ранних стадиях болезни, еще до проявления симптомов заболевания, а также структурных и анатомических изменений в органах организма.

Преимущество ПЭТ диагностики также состоит в получении наиболее точной оценки степени метастазирования онкологических заболеваний в сравнении с оценками, полученными с помощью рентгеновской компьютерной или магнитно-резонансной томографии [1].

Помимо диагностики раковых заболеваний, ПЭТ находит применение в кардиологии, неврологии, эндокринологии и других областях медицины.

В ПЭТ-исследовании позитрон-излучающий радиофармпрепарат (меченое радионуклидом вещество) вводится пациенту внутривенно. После чего переносится кровью и достигает области интереса, например, структур головного мозга или сердечной мышцы. При столкновении позитрона и свободного электрона области, происходит их аннигиляция и образование двух фотонов с энергией 511 кэВ, разлетающихся в противоположных направлениях. Система детектирования томографа регистрирует аннигиляционные гамма-кванты и определяет точки аннигиляции, находящиеся на одной линии совпадения. Далее система суммирует все линии совпадения, зарегистрированные за время экспозиции, и производит реконструкцию данных. Таким образом, после реконструкции получают распределение позитрон-излучающего радиофармпрепарата (РФП) [2].

Одним из методов математической обработки данных в ПЭТ является метод камерных моделей, который позволяет анализировать распределение РФП во времени, используя дифференциальные уравнения первого порядка с постоянными коэффициентами.

В рамках выпускной квалификационной работы рассматриваются камерные модели в исследованиях мозга человека, проводится математическая обработка данных ПЭТ исследований.

Постановка задачи

В исследованиях головного мозга человека физиологически значимым параметром является потенциал связи (Binding Potential – BP). Этот параметр отражает концентрацию(плотность) рецепторов в области интереса мозга [3].

Целью данной работы является изучить ПЭТ исследования головного мозга человека, рассмотреть метод камерных моделей, который находит применение в данной области, получить наиболее эффективную оценку физиологически-значимого параметра – потенциала связи для построенной модели исследований.

В данной работе оценка потенциала связи проводится на основе камерного моделирования с использованием метода Ichise. Полученная регрессионная модель является чувствительной к статистическому шуму, что происходит в силу наличия ошибок в зависимой и независимых переменных [4]. В итоге оценка параметра методом наименьших квадратов (который в большинстве случаев используется для получения оценок), оказывается смещённой и несостоятельной [9]. В качестве регрессионной модели, которая бы учитывала данные особенности, предлагается рассмотреть метод регрессии на главные компоненты, а также метод гребневой регрессии с использованием кросс-валидации по K блокам.

Глава 1. Модель исследования

1.1. Камерные модели в исследованиях головного мозга

В исследованиях головного мозга используются двух- и трех-камерные модели.

РФП вводится внутривенно, затем попадает в сердце и далее в артериальную кровь. Артериальная кровь является первой камерой (C_p). Далее из артериальной крови РФП проходит гемато-энцефалический барьер и попадает во вторую камеру (C_2, C'_2), которая состоит из множества областей, включая внутритканевые жидкости и внутриклеточную цитоплазму. Третья камера (C_3) – является областью специфического «связывания» и содержит высокоаффинные рецепторы, т.е. рецепторы, которые обладают высокой степенью сродства к индикатору. Области мозга, в которых такие рецепторы отсутствуют, относятся к референтным областям и не содержат третьей камеры, что соответствует двухкамерной модели. Кроме того, существует камера неспецифического «связывания». На практике эта камера находится в быстром равновесии со второй камерой и вместе две эти камеры рассматриваются как одна.

Стоит отметить, что в системе человеческого организма происходит постоянное очищение первой камеры. Более того, перемещение РФП между камерами зависит от регионального мозгового кровообращения.

В камерном анализе вводятся следующие физиологически обоснованные предположения:

1. РФП в систему поступает из единственного источника – артериальной плазмы;
2. РФП может свободно проникать между первой и второй камерами и обратно;
3. рассматриваются реакции только первого порядка, все реакции происходят мгновенно. Это даёт нам основания применять

дифференциальные уравнения первого порядка с постоянными коэффициентами;

4. объемы распределения (отношение концентрации РФП в рассматриваемой области к концентрации РФП в плазме крови в момент равновесия) во вторых камерах референтной (V_2') и рецепторной (V_2) областей предполагаются идентичными.

Таким образом, камерная модель мозга человека состоит из камеры артериальной крови (C_p), внутримозговой камеры (C_2) и камеры специфического связывания (C_3). Аналогично C_p и C_2' для референтных областей мозга (рис. 1).

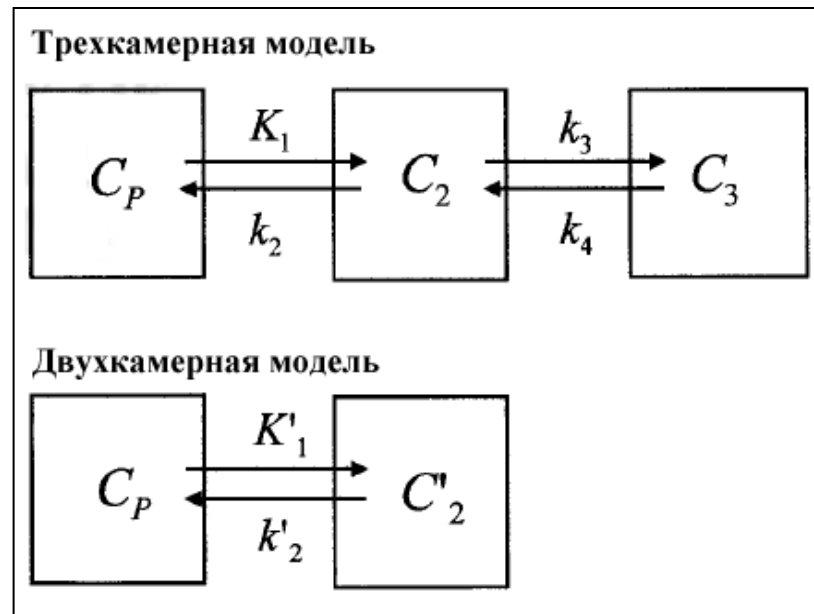


Рис. 1. Трехкамерная и двухкамерная модели, используемые для мозга человека.

Изменение концентрации РФП в камерах описывается следующей системой дифференциальных уравнений:

$$\frac{dC_3(t)}{dt} = k_3 C_2(t) - k_4 C_3(t), \quad (1)$$

$$\frac{dC_2(t)}{dt} = K_1 C_1(t) - k_2 C_2(t) - k_3 C_2(t) + k_4 C_3(t), \quad (2)$$

$$\frac{dC'_2(t)}{dt} = K'_1 C_1(t) - k'_2 C_2(t), \quad (3)$$

где $K_1, K'_1, k_2, k'_2, k_3, k_4$ — константы транспорта, $C_p(t)$ — суммарная концентрация РФП в плазме, f_1 — фракция активности неметаболизированного

РФП в плазме, тогда $C_1(t) = f_1 C_p(t)$.

Полученная при сканировании информация используется для нахождения параметров в состоянии равновесия, стоит отметить, что на самом деле равновесия не достигается, однако такой подход обусловлен методом. Левые части уравнений (1) - (3) будут равны нулю, т.к. не происходит перемещения РФП между камерами. Тогда из уравнения (1) получим (4):

$$\frac{C_3}{C_2} = \frac{k_3}{k_4}. \quad (4)$$

Выражая C_2 из уравнения (4) и подставляя в уравнение (2), получим (5):

$$\frac{C_3}{f_1 C_p} = \frac{K_1 k_3}{k_2 k_4}. \quad (5)$$

Из уравнения (3) получим уравнение (6):

$$\frac{C'_2}{f_1 C_p} = \frac{K'_1}{k'_2}. \quad (6)$$

Количественная оценка РФП вводится как объем распределения, который определяется как отношение концентрации РФП в рассматриваемой области к концентрации РФП в плазме крови в момент равновесия. Тогда объем распределения в областях может быть описан уравнениями (7) - (9):

$$V_3 = \frac{C_3}{f_1 C_p} = \frac{K_1 k_3}{k_2 k_4}, \quad (7)$$

$$V_2 = \frac{C_2}{f_1 C_p} = \frac{K_1}{k_2}, \quad (8)$$

$$V'_2 = \frac{C'_2}{f_1 C_p} = \frac{K'_1}{k'_2}. \quad (9)$$

Таким образом, потенциал связи в исследованиях головного мозга определяется уравнением (10):

$$BP = \frac{C_3}{f_1 C_p} = V_3. \quad (10)$$

Стоит отметить, что в данном определении потенциала связи требуется периодический забор крови. При использовании модели Ichise, которая не требует заборов крови и будет описана далее, потенциал связи определяется

как отношение V_3 к V_2 (обозначается как потенциал связи BP^*) (11) [3]:

$$BP^* = \frac{V_3}{V_2} = \frac{BP}{V_2} = \frac{C_3/f_1 C_p}{C_2/f_1 C_p} = \frac{C_3}{C_2} = \frac{k_3}{k_4} \quad (11)$$

1.2. Графическая модель Logan

Одним из наиболее широко используемых РФП для исследований мозга в ПЭТ является [^{11}C]Raclopride, который связывается с дофаминовым D_2 рецептором [5, 6]. Для данного РФП референтной областью служит мозжечок (Cerebellum), а область интереса представляет собой полосатое тело (Striatum) головного мозга [7].

После оконтуривания областей ПЭТ-изображений получают входные данные как графики зависимости активности в областях от времени (рис. 3):

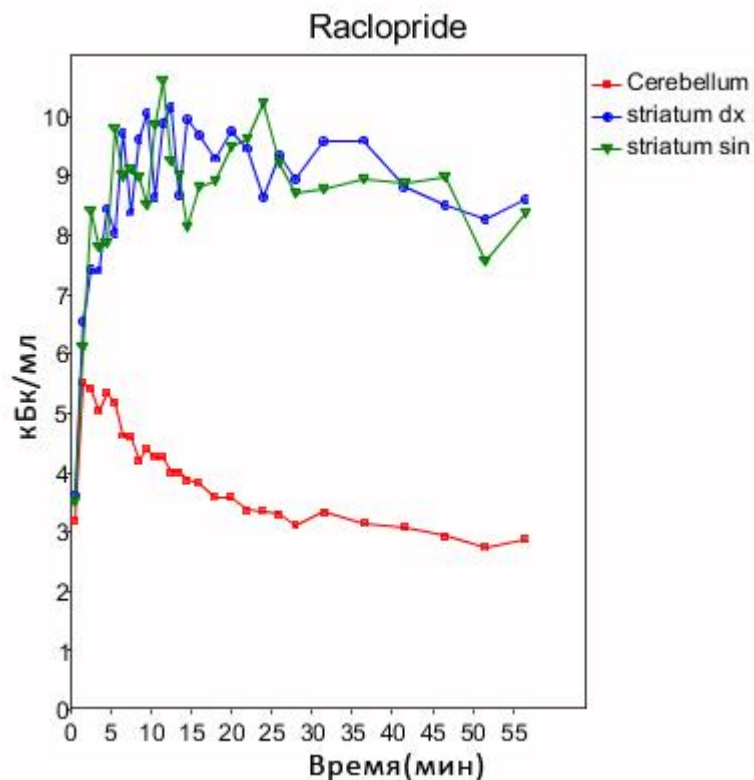


Рис. 3. Графики зависимости активности РФП в области интереса (в левой и правой части полосатого тела (Striatum dx и Striatum sin)) и референтной области (Cerebellum) от времени [8].

Активность РФП может быть представлена выражением (12) для области интереса и выражением (13) для референтной области:

$$C_{RC}(t) = C_2(t) + C_3(t) + C_p(t)V_1, \quad (12)$$

$$C_{RF}(t) = C'_2(t) + C_p(t)V_1, \quad (13)$$

где V_1 — объем сосудов внутри рассматриваемой области. V_1 оценивают в 5% и он не оказывает значительного вклада в исследование.

Метод графического определения потенциала связи Logan состоит в том, что графики концентрации РФП в рассматриваемой области и в артериальной крови от времени объединяются в один так, что при определенных условиях этот график становится линейным. Его наклон равен объему распределения РФП в рассматриваемой области.

Для двухкамерной модели интегрирование обеих частей уравнения (3) дает (14):

$$k'_2 \int_0^t C'_2(t) dt = K'_1 f_1 \int_0^t C_p(t) dt - C'_2(t). \quad (14)$$

Используя подстановку $C'_2(t) = C_p(t)V_1 - C_{RF}$ из уравнения (13) и то, что $V'_2 = K'_1/k'_2$, и разделив выражение (14) на $C_{RF}(t)$, получим уравнение (15):

$$\frac{\int_0^t C_{RF}(t) dt}{C_{RF}(t)} = (V'_2 f_1 + V_1) \frac{\int_0^t C_p(t) dt}{C_{RF}(t)} + \frac{1}{k'_2} \left(\frac{V_1 C_p(t)}{C_{RF}(t)} - 1 \right). \quad (15)$$

Спустя некоторое время график $\int_0^t C_{RF}(t) dt / C_{RF}(t)$ от $\int_0^t C_p(t) dt / C_{RF}(t)$ становится линейным с наклоном $a' = V'_2 f_1$ и остаточным членом $b' = -\frac{1}{k'_2}$.

В силу того, что $V'_2 \gg V_1$ и вклад $V_1 C_p(t) / C_{RF}(t)$ мал, имеем (16):

$$\frac{\int_0^t C_{RF}(t) dt}{C_{RF}(t)} = V'_2 f_1 \frac{\int_0^t C_p(t) dt}{C_{RF}(t)} + \frac{1}{k'_2}. \quad (16)$$

Для трехкамерной модели аналогичное выражение получается из преобразований уравнений (1) и (2). Исследования показали, что спустя некоторое время последний член в уравнении (17) становится константой, если выполняются условия: $C_2(t)/C_p(t)$ и $C_3(t)/C_p(t)$ постоянны.

$$\frac{\int_0^t C_{RC}(t) dt}{C_{RC}(t)} = (V_2 f_1 + f_1 V_3 + V_1) \frac{\int_0^t C_p(t) dt}{C_{RC}(t)} + b. \quad (17)$$

Тогда график $\int_0^t C_{RC}(t) dt / C_{RC}(t)$ от $\int_0^t C_p(t) dt / C_{RC}(t)$ имеет наклон $a = V_2 f_1 + f_1 V_3 + V_1$ и остаток b . Если предположить, что $V'_2 = V_2$ и пренебречь V_1 , то выражение для потенциала связи принимает вид:

$$BP = \frac{1}{f_1} (a - a') \quad (18)$$

ИЛИ

$$BP^* = \frac{a}{a'} - 1. \quad (19)$$

1.3. Неинвазивная модель Ichise

Рассмотренный выше графический метод Logan определения потенциала связи требует периодического взятия проб артериальной крови, что также является источником статистических ошибок и причиняет дискомфорт пациентам.

Метод, предложенный Ichise, является следствием из уравнений, описанных выше.

Выражая из уравнения (15) интеграл $\int_0^t C_p(t)dt$ и подставляя выражение для него в уравнение (17), получим (20):

$$\frac{\int_0^t C_{RC}(t)dt}{C_{RC}(t)} = \left(\frac{a}{a'}\right) \frac{\int_0^t C_{RF}(t)dt}{C_{RC}(t)} + \left(-\frac{ab'}{a'}\right) \frac{C_{RF}(t)}{C_{RC}(t)} + b. \quad (20)$$

В итоге, потенциал связи

$$BP^* = \frac{a}{a'} - 1$$

может быть оценен методами множественного регрессионного анализа. В отличие от графического метода, предложенного Logan, данный метод не требует значения k'_2 (т.е. заборов крови во время проведения исследования) [3].

Глава 2. Математическая обработка данных

2.1. Модель наблюдения

В соответствии с уравнением (20) введем следующие обозначения для векторов данных, необходимые для дальнейшего анализа (21) - (23):

$$\frac{\int_0^t C_{RC}(t)dt}{C_{RC}(t)} = Y^*, \quad (21)$$

$$\frac{\int_0^t C_{RF}(t)dt}{C_{RC}(t)} = X^*, \quad (22)$$

$$\frac{C_{RF}(t)}{C_{RC}(t)} = Z^*. \quad (23)$$

Также в соответствии с уравнением (20) обозначим коэффициенты:

$$b = \beta_1, \left(\frac{a}{a'}\right) = \beta_2, \left(-\frac{ab'}{a'}\right) = \beta_3. \quad (24)$$

Таким образом, получим следующую модель:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i, \quad (25)$$

где Y_i, X_i, Z_i – точные значения измеряемых величин.

Пусть U_i, V_i, W_i – ошибки, которые вводятся вследствие неточности измерений и вычисления интегралов, тогда:

$$\begin{aligned} Y_i^* &= Y_i + U_i, \\ X_i^* &= X_i + V_i, \\ Z_i^* &= Z_i + W_i. \end{aligned} \quad (26)$$

Из уравнений (20) – (26) модель наблюдений представляется в виде (27):

$$Y_i^* = \beta_1 + \beta_2 X_i^* + \beta_3 Z_i^* + U_i - \beta_2 V_i - \beta_3 W_i. \quad (27)$$

Из выражения (27) следует, что параметры β_2 и β_3 содержатся в $\varepsilon_i = U_i - \beta_2 V_i - \beta_3 W_i$. В этом случае ε_i имеет корреляции с ошибками V_i и W_i независимых переменных X_i и Z_i соответственно.

Корреляции между независимыми переменными и ошибками приводят к смещенным и несостоятельным оценкам, которые получаются при использовании метода наименьших квадратов [9].

Таким образом, задача сводится к определению эффективной оценки коэффициента β_2 в наблюдаемой модели (27), т.к. именно этот коэффициент определяет искомое значение потенциала связи BP^* .

2.2. Регрессия на главные компоненты

Идея метода главных компонент состоит во введении новых искусственных переменных W_j (28), которые являются линейными комбинациями исходных переменных, причем новые переменные выбираются таким образом, чтобы полученные вектор столбцы элементов W_{ij} были ортогональны.

$$W_j = \gamma_{1j}S_1 + \gamma_{2j}S_2, \dots, \gamma_{rj}S_r. \quad (28)$$

Рассмотрим матрицу независимых переменных S , в столбцах которой находятся значения исходных переменных, центрируем ее, т.е.:

$$s_{ij} = S_{ji} - \bar{S}_j, \quad (29)$$

$$\bar{S}_j = \frac{\sum_{i=1}^n S_{ji}}{n}. \quad (30)$$

Для того чтобы найти главные компоненты и коэффициенты, с которыми в главные компоненты входят элементы матрицы S , будем использовать сингулярное (SVD) разложение [12, 13].

Для вещественной матрицы S существуют две вещественные ортогональные матрицы Q и P такие, что $Q^T S P$ – диагональная матрица H :

$$H = Q^T S P. \quad (31)$$

Матрицы Q и P выбираются так, чтобы диагональные элементы матрицы H (сингулярные числа) имели вид $\sigma_1 \geq \sigma_2 \dots \geq \sigma_r > 0$, где r – ранг матрицы S . Эквивалентная запись сингулярного разложения:

$$S = Q H P^T. \quad (32)$$

Тогда произведение матриц QH является матрицей, содержащей значения главных компонент, а матрица P отвечает коэффициентам (весам), с которыми исходные переменные входят новые главные компоненты (28).

Первая главная компонента W_j соответствует наибольшему собственному числу матрицы $S^T S$ (квадрату сингулярного числа) и «объясняет» наибольшую часть вариации данных. Полученные далее новые переменные объясняют все меньшие доли вариации.

Обычно используют только часть новых переменных, например, пока эти переменные не станут «объяснять» большую часть суммарной дисперсии (75%) [10].

После выделения главных компонент используется метод наименьших квадратов для получения уравнения, связывающего в нашем случае Y^* с выбранными главными компонентами. В силу того что уравнение будет получено в виде функции от компонент W_j , его можно преобразовать и выразить через исходные вектора X^* и Z^* , тем самым получим оценку коэффициента β_2 .

2.3. Гребневая регрессия и кросс-валидация

Метод гребневой регрессии реализуется с помощью введения дополнительного слагаемого в исходный минимизируемый функционал.

Будем так же рассматривать матрицу независимых переменных S . Если матрица $S^T S$ является плохо обусловленной (почти вырожденной), это приводит к неустойчивости решения. Параметры модели находятся из минимизации функционала (33):

$$\beta^* = \arg \min_{\beta} \left(\sum_{i=1}^m \left(y_i - \sum_{j=1}^n \beta^j s_j^i \right)^2 + \tau \|\beta\|^2 \right), \quad (33)$$

что эквивалентно (34):

$$\beta^* = \arg \min_{\beta} \|Y^* - S\beta\|^2, \|\beta\|^2 \leq C. \quad (34)$$

Решением задачи минимизации является вектор (35):

$$\beta^* = (S^T S + \tau I_n)^{-1} S^T Y^* \quad (35)$$

Рассмотрим сингулярное разложение $S = QHP^T$. Числом обусловленности матрицы называется отношение максимального сингулярного числа к минимальному (36):

$$k = \frac{\sigma_1}{\sigma_r}. \quad (36)$$

Число обусловленности k регуляризованной матрицы S (37):

$$k(S^T S + \tau I) = \frac{\sigma_1 + \tau}{\sigma_r + \tau}. \quad (37)$$

С увеличением параметра регуляризации τ уменьшается число обусловленности матрицы $S^T S$ и повышается устойчивость решения нормального уравнения регрессии [14].

В настоящее время не существует определенного подхода, регулирующего выбор параметра τ в каждом отдельном случае. На параметр накладывается ряд ограничений, которые позволяют выбрать подходящее значение τ [10, 11].

Одним из методов определения параметра регуляризации является получение наименьшего среднего квадрата ошибки MSE , соответствующего параметру τ :

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i^*)^2, \quad (38)$$

где \hat{Y}_i – предсказанные значения Y_i^* .

Для определения наименьшей оценки MSE воспользуемся методом кросс-валидации (перекрестной проверкой) по K блокам.

В кросс-валидации по K блокам исходный набор данных разбивается на K одинаковых по размеру блока. Причем один из K блоков оставляется для тестирования модели, а оставшиеся $K - 1$ блоки используются как тренировочный набор. Процесс повторяется K раз, т.е. каждый из блоков является тестовым набором один раз. В итоге имеем K результатов, соответствующих каждому блоку, далее они усредняются. Получается, что все наблюдения используются и для тренировки, и для тестирования модели. [15, 16].

Таким образом, каждый раз предсказывая вектор значений \hat{Y} , применяя метод кросс-валидации по K блокам, можно вычислить наименьшую оценку MSE для каждого параметра τ из заданного диапазона. Интересующие нас коэффициенты β , соответствующие модели с наименьшим MSE , будут иметь смещение, однако такая оценка оказывается более эффективной, чем оценка по методу наименьших квадратов [10].

2.4. Применение методов и анализ результатов

В данной работе эксперимент проводился на тестовых данных [^{11}C]Raclopride [8]. Аналогично формулам (21) - (23) были сформированы вектора независимых переменных X и Z , а также значения вектора Y . Выбраны коэффициенты $B_1 = -10$, $B_2 = 2,7$, $B_3 = 11$, относительно которых были получены тестовые значения вектора Y , которые мы далее принимаем за истинные.

Вектора ошибок U и W для Y и Z соответственно формировались с помощью генерации случайных чисел. Эти вектора не изменяются в ходе эксперимента и подчиняются закону нормального распределения со стандартными отклонениями $\sigma_y = 0,45$ и $\sigma_z = 0,05$, которые выбирались из условий (39):

$$\begin{aligned} 2\sigma_y &< \min|Y_{i+1} - Y_i|, \\ 2\sigma_z &< \min|Z_{i+1} - Z_i|. \end{aligned} \tag{39}$$

Ошибки вектора X подчиняются стандартному нормальному распределению. Проводится 100 опытов, где каждый раз формируется вектор ошибок V для X и далее находится оценка коэффициента β_2 модели наблюдений регрессией на главные компоненты, гребневой регрессией с использованием кросс-валидации и методом наименьших квадратов.

Эксперимент был проведен в программном пакете Matlab (см. Приложение), результаты вычислений коэффициента β_2 для разных регрессионных методов приведены в Таблице 1.

Поскольку оценка методом регрессии на главные компоненты в случае совпадения количества независимых переменных и главных компонент дает в точности такую же оценку, как и метод наименьших квадратов, регрессия осуществляется только на первую главную компоненту, которая «объясняет» наибольшую часть вариации исходных данных. Для того чтобы в этом убедиться, в Таблице 1 приводится значение наибольшего собственного числа λ_r нормированной матрицы $S^T S$, которое в этом случае есть не что иное, как

дисперсия первой главной компоненты (переведено в проценты). Сумма всех λ_r равна полной (суммарной) дисперсии переменных [10].

Сравнение оценок исследуемого коэффициента β_2 приведено на диаграмме (рис.4).

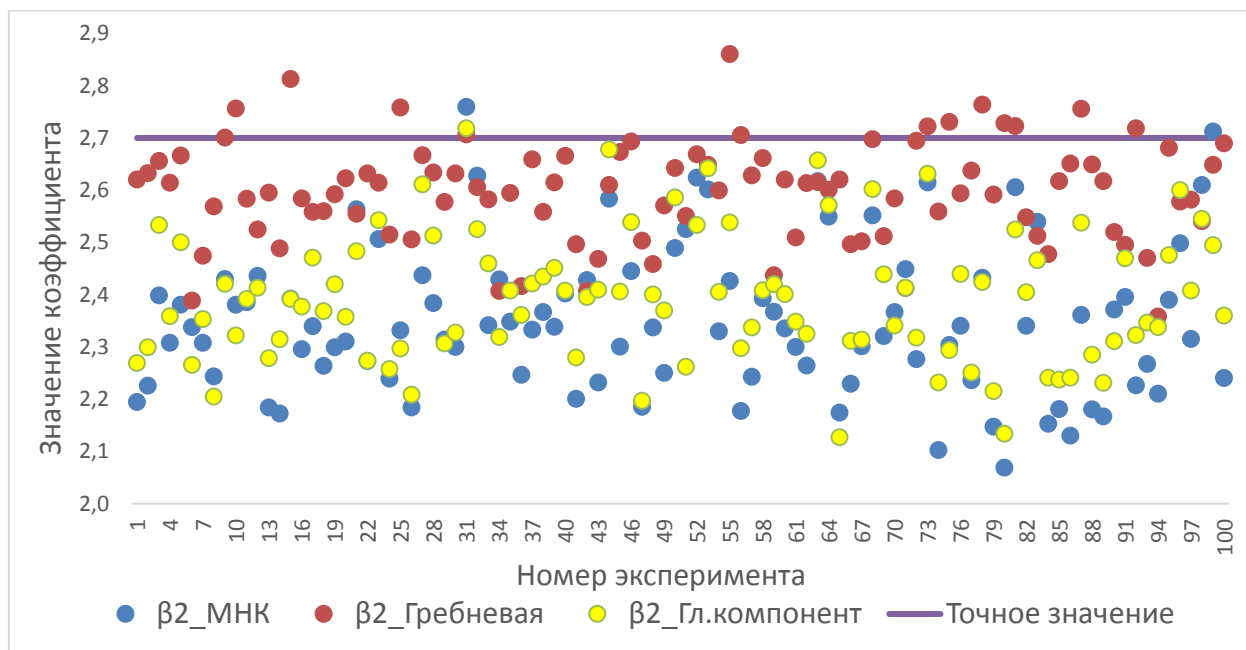


Рис. 4. Сравнение оценок β_2 , полученных в результате эксперимента.

Таким образом, оценка, полученная методом гребневой регрессии, оказывается более точной, чем оценка МНК, в 91% случаев. Оценка методом регрессии на главные компоненты оказывается более точной, чем оценка МНК, в 76% случаев. В сумме относительно регрессии на главные компоненты и метода наименьших квадратов оценка с применением гребневой регрессии является наиболее эффективной в 88% случаев.

Таблица 1.

№	$\beta_{2_МНК}$	$\beta_{2_Гребн}$	$\beta_{2_Глав}$	$\lambda_r, \%$	№	$\beta_{2_МНК}$	$\beta_{2_Гребн}$	$\beta_{2_Глав}$	$\lambda_r, \%$
1	2,1949	2,6205	2,2699	76,53	51	2,5259	2,5509	2,2624	83,58
2	2,2259	2,6328	2,2996	77,84	52	2,6242	2,6690	2,5335	80,33
3	2,3991	2,6563	2,5337	74,76	53	2,6015	2,6489	2,6423	77,98
4	2,3081	2,6147	2,3594	77,05	54	2,3306	2,5997	2,4056	76,82
5	2,3814	2,6668	2,5009	74,81	55	2,4267	2,8613	2,5388	76,09
6	2,3382	2,3890	2,2661	80,91	56	2,1778	2,7057	2,2979	74,76
7	2,3080	2,4748	2,3538	78,00	57	2,2434	2,6290	2,3374	76,36
8	2,2437	2,5693	2,2052	79,87	58	2,3933	2,6616	2,4090	78,48
9	2,4302	2,7010	2,4212	79,09	59	2,3677	2,4376	2,4206	78,38
10	2,3809	2,7566	2,3223	80,15	60	2,3357	2,6208	2,4014	76,76

11	2,3860	2,5842	2,3919	78,59	61	2,3002	2,5100	2,3489	77,34
12	2,4362	2,5250	2,4137	80,06	62	2,2647	2,6138	2,3254	76,84
13	2,1842	2,5957	2,2787	75,56	63	2,6178	2,6157	2,6578	78,01
14	2,1730	2,4892	2,3147	74,27	64	2,5500	2,6014	2,5720	78,05
15	2,3927	2,8131	2,3921	78,88	65	2,1745	2,6208	2,1276	79,98
16	2,2960	2,5847	2,3776	77,06	66	2,2297	2,4973	2,3122	76,78
17	2,3400	2,5587	2,4715	74,54	67	2,3016	2,5028	2,3145	78,78
18	2,2643	2,5601	2,3690	76,29	68	2,5519	2,6981	2,6027	77,38
19	2,2999	2,5923	2,4203	75,49	69	2,3214	2,5124	2,4392	76,07
20	2,3110	2,6231	2,3579	77,26	70	2,3672	2,5846	2,3415	79,71
21	2,5638	2,5551	2,4835	80,83	71	2,4492	2,4133	2,4134	79,51
22	2,2732	2,6323	2,2741	78,87	72	2,2767	2,6947	2,3183	78,07
23	2,5067	2,6147	2,5432	77,54	73	2,6155	2,7225	2,6320	78,50
24	2,2398	2,5155	2,2577	77,93	74	2,1028	2,5597	2,2325	74,13
25	2,3319	2,7584	2,2974	79,57	75	2,3042	2,7311	2,2942	78,85
26	2,1845	2,5059	2,2091	78,50	76	2,3405	2,5946	2,4402	76,03
27	2,4378	2,6675	2,6115	74,15	77	2,2368	2,6381	2,2512	78,73
28	2,3845	2,6342	2,5137	75,56	78	2,4328	2,7643	2,4247	79,68
29	2,3145	2,5781	2,3069	79,46	79	2,1476	2,5922	2,2158	76,65
30	2,3003	2,6324	2,3276	78,48	80	2,0693	2,7285	2,1340	76,97
31	2,7600	2,7075	2,7189	80,03	81	2,6060	2,7231	2,5250	80,75
32	2,6284	2,6061	2,5260	80,73	82	2,3408	2,5483	2,4052	77,27
33	2,3419	2,5825	2,4604	75,55	83	2,5402	2,5130	2,4665	80,47
34	2,4297	2,4076	2,3195	80,99	84	2,1531	2,4779	2,2416	71,34
35	2,3485	2,5953	2,4083	77,41	85	2,1811	2,6179	2,2375	77,88
36	2,2471	2,4165	2,3617	75,83	86	2,1307	2,6516	2,2412	75,19
37	2,3332	2,6597	2,4218	77,20	87	2,3616	2,7560	2,5381	73,05
38	2,3669	2,5592	2,4349	76,79	88	2,1809	2,6495	2,2855	75,64
39	2,3387	2,6151	2,4519	75,10	89	2,1675	2,6178	2,2315	77,17
40	2,4027	2,6659	2,4084	79,11	90	2,3721	2,5200	2,3112	80,17
41	2,2009	2,4968	2,2801	76,48	91	2,3958	2,4960	2,4701	77,20
42	2,4282	2,4078	2,3959	80,55	92	2,2266	2,7186	2,3228	77,42
43	2,2320	2,4691	2,4098	73,62	93	2,2676	2,4704	2,3467	77,17
44	2,5843	2,6105	2,6782	76,59	94	2,2110	2,3586	2,3383	74,82
45	2,3008	2,6737	2,4065	76,25	95	2,3905	2,6812	2,4759	76,49
46	2,4454	2,6937	2,5395	76,86	96	2,4991	2,5783	2,6005	76,28
47	2,1857	2,5038	2,1974	78,86	97	2,3154	2,5820	2,4080	76,47
48	2,3379	2,4587	2,4010	78,20	98	2,6102	2,5409	2,5452	80,44
49	2,2507	2,5713	2,3699	75,40	99	2,7128	2,6488	2,4953	82,60
50	2,4897	2,6427	2,5865	76,12	100	2,2406	2,6899	2,3606	75,44

Среднее процентное соотношение дисперсии λ_r в методе регрессии на главные компоненты составляет 78%, именно такую часть вариации исходных данных в среднем «объясняет» первая главная компонента.

Заключение

В рамках данной работы были изучены ПЭТ исследования головного мозга человека. Рассмотрен, применимый в данной области медицины, метод камерных моделей. Выбрана модель исследования и получена оценка физиологически-значимого параметра – потенциала связи, с применением различных регрессионных моделей. Полученные результаты могут быть полезны при обработке данных ПЭТ исследований головного мозга, а также для дальнейших исследований в данной области медицины.

Список литературы

1. Беляев В.Н., Климанов В.А. Физика ядерной медицины. Ч.1. Учебное пособие. – М.: НИЯУ МИФИ, 2012. 308 с.
2. Котина Е.Д. Некоторые вопросы моделирования динамических процессов в радионуклидных исследованиях. – СПб.: Изд-во ВВМ, 2013. 150 с.
3. Ichise M., Meyer J.H., Yonekura Y. An introduction to PET and SPECT neuroreceptor quantification models. *J Nucl Med*, 2001; 42:755-763.
4. Ichise M., Toyama H., Innis R.B., Carson R.E. Strategies to improve neuroreceptor parameter estimation by linear regression analysis. *J Cereb Blood Flow Metab*, 2002. 22:1271-1281.
5. Farde L., Ehrin E., Eriksson L., Greitz T., Hall H., Hedström C.G., Litton J.E., Sedvall G. Substituted benzamides as ligands for visualization of dopamine receptor binding in the human brain by positron emission tomography. *Proc Natl Acad Sci USA* 1985; 82: 3863-3867.
6. Laruelle M. Imaging synaptic neurotransmission with in vivo binding competition techniques: a critical review. *J Cereb Blood Flow Metab*. 2000; 20: 423-451.
7. Quantification of [¹¹C]Raclopride PET studies.
http://www.turkupetcentre.net/petanalysis/analysis_11c-raclopride.html.
8. Regional PET Time-Activity Curve.
http://www.turkupetcentre.net/petanalysis/tac_regional.html.
9. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. – М.: Дело, 2004. 320 с.
10. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Т.2. – М.: Финансы и статистика, 1987. 351 с.
11. Демиденко Е. З. Линейная и нелинейная регрессии. – М.: Финансы и статистика, 1981. 302 с.
12. Стренг Г. Линейная алгебра и ее применения. – М.: Мир, 1980. 456 с.
13. Голуб Дж., Ван-Лоун Ч. Матричные вычисления. – М.: Мир. 1999. 548 с.
14. Стрижов В.В., Крымова Е.А. Методы выбора регрессионных моделей. –

М.: ВЦ РАН, 2010. 60 с.

15. Воронцов К. В. Комбинаторный подход к оценке качества обучаемых алгоритмов. — Математические вопросы кибернетики / Под ред. О. Б.

Лупанов. — М.: Физматлит, 2004. — Т. 13. С. 5–36.

16. Kohavi R. A. Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. — 14th International Joint Conference on Artificial Intelligence, Palais de Congres Montreal, Quebec, Canada, 1995. С. 1137-1145.

Приложение

```
% Вектор X без ошибки, вектора Y* и Z* с фиксированной ошибкой
X =
[1.3589;2.1327;2.7708;2.762;3.5515;4.0172;4.5695;5.3173;5.0281;5
.0779;6.2751;6.8815;8.0967;8.3473;9.0705;9.2933;9.8918;9.9694;11
.7555;13.2026;14.9255;16.4561;18.3029;19.8063;25.4133;24.9359;];
Y = [-
0.0492;2.9900;5.1810;2.5353;4.8409;6.6693;7.0576;9.4794;8.4886;9
.0991;12.3031;12.5818;15.4543;16.8494;19.8626;19.1207;21.4611;20
.7695;24.4920;31.0235;34.2045;38.4661;43.7515;47.2168;60.4261;60
.7800;];
Z =
[0.6323;0.7007;0.6230;0.5295;0.5411;0.5601;0.5471;0.5197;0.3574;
0.3637;0.3803;0.5624;0.4445;0.4701;0.3910;0.4229;0.3117;0.2575;0
.2867;0.3826;0.3709;0.3419;0.4183;0.3405;0.3715;0.4238;];

% Формируем ошибку вектора X
V = randn(26,1);
X = X+V;

% Центрируем вектора исходных данных
for i = 1:26
    X_c(i) = X(i)-mean(X);
    Z_c(i) = Z(i)-mean(Z);
end

M = 0; N = 0;
for i= 1:26
    N = N + (X(i)-mean(X))^2;
    M = M + (Z(i)-mean(Z))^2;
end

% Нормируем вектора исходных данных
for i = 1:26
    X_norm(i) = X_c(i)/sqrt(N);
    Z_norm(i) = Z_c(i)/sqrt(M);
end

% Будем использовать центрированную матрицу S_c = [X* Z*] для
Метода регрессии на главные компоненты, т.к. исходные вектора не
имеют сильного отличия в порядках значений
S_c = [X_c' Z_c'];

% Формируем матрицу S_mnk для Метода наименьших квадратов
S_mnk = [ones(26,1) X Z];

% Формируем матрицу S_gr для Метода гребневой регрессии
S_gr = [X Z];
```

```

% Сингулярное разложение матрицы S
[Q, H, P] = svd(S_c);

% Формируем матрицу, содержащую значения главных компонент
R = Q*H;

% Формируем нормированную матрицу S_norm
S_norm = [X_norm' Z_norm'];

% Вычисляем собственные числа матрицы S_norm'S_norm
K = eig(S_norm'*S_norm);

% Вычисляем процент дисперсии исходных данных, которая объясняет
первая главная компонента
disp = max(K)*100/sum(K);

% Регрессия на первую главную компоненту
f = regress(Y, R(:,1));
% Оценка коэффициента B2_comp
B2_comp = P(1,1)*f;

% Регрессия по методу наименьших квадратов
mnk = regress(Y, S_mnk);
% Оценка коэффициента B2_mnk
B2_mnk = mnk(2);

% Кросс-валидация с разбиением на 10 блоков
k = 0:.5:200;
MSE = zeros(size(k));
for i = 1:length(k)
    regf = @(XTRAIN,ytrain,XTEST) (XTEST*ridge(ytrain, XTRAIN,
k(i), 1));
    MSE(i) = crossval('mse', S_gr, Y, 'kfold', 10, 'predfun',
regf);
end

% Вычисление наименьшего MSE
[MSE I] = min(MSE);
% Гребневая регрессия
G = ridge(Y,S_gr,k(I));
% Оценка коэффициента B2_ridge
B2_ridge = G(1);

```